Abduction aiming at empirical progress or even at truth approximation, leading to challenge for computational modelling

Kuipers, Theo A.F.

Published in:
Scientific Discovery An Creativity, vol. 4 (Foudation), 307-323

THEO A.F. KUIPERS

ABDUCTION AIMING AT EMPIRICAL PROGRESS OR EVEN TRUTH APPROXIMATION LEADING TO A CHALLENGE FOR COMPUTATIONAL MODELLING

ABSTRACT. This paper primarily deals with the conceptual prospects for generalizing the aim of abduction from the standard one of explaining surprising or anomalous observations to that of empirical progress or even truth approximation. It turns out that the main abduction task then becomes the instrumentalist task of theory revision aiming at an empirically more successful theory, relative to the available data, but not necessarily compatible with them. The rest, that is, genuine empirical progress as well as observational, referential and theoretical truth approximation, is a matter of evaluation and selection, and possibly new generation tasks for further improvement. The paper concludes with a survey of possible points of departure, in AI and logic, for computational treatment of the instrumentalist task guided by the 'comparative evaluation matrix'.

KEY WORDS: abduction, empirical progress, truth approximation, truthlikeness

1. INTRODUCTION

In the literature there is more or less agreement about the general nature of abduction. Following Aliseda (1997), it amounts to *the search for an acceptable explanatory hypothesis for a surprising or anomalous (individual or general) observational fact*. To get a sharper idea of it, we need to make two distinctions. First, there is the distinction between the *process* of abduction and the *product*. Second, the process can be subdivided into three phases: the *generation* of one or more hypotheses, their *evaluation*, and the *selection* of one of them as the best one. The second subdivision may be merely analytical, for the three aspects of the process may in fact be integrated. Combining the two distinctions, we get the following division of the total product in kinds of subproducts: the generated hypotheses, the evaluation reports of them, and the selected hypothesis. There has been a lot of dispute about whether Peirce meant to

Foundations of Science **4:** 307–323, 1999.

include the evaluation and selection phases in his notion of abduction, and his preference in this respect seems to have changed. Neglecting this exegetical point further, we propose to speak only of (the process of) abduction when the generation, evaluation, and selection of hypotheses are all included, and to speak of abduction *in the strict sense* when only the generation is meant.

Our investigations of types of scientific processes or products are governed by the working hypothesis that the analysis of a type of product may suggest an analysis of the process, and vice versa. For example, in her discussion of abduction aiming at the explanation of a surprising fact, Aliseda starts from the following general characterization of the (total) product: a hypothesis H, such that B, H $\Rightarrow$ E, where B represents the background knowledge, H the generated hypothesis, E the surprising fact to be explained, and '$\Rightarrow$' a formal relation which has to be specified further, e.g. as deductive consequence. In line with this product searched for, Aliseda's hypothesis generation algorithm systematically seeks, for given B and E, hypotheses that close the deductive(ly specified) gap between B and E. In the following we will use the same type of product/process interaction.

From the general characterization in the beginning of this section, it is clear that there are at least two more specific abduction tasks, where we limit ourselves to the standard, 'deductive' versions.

Task I. In the case of a surprising observation, i.e. the situation that E is not entailed by B, the task is to expand B with some H such that B&H entails E, but H alone does not. This is called *novelty guided abduction*.

Task II. In the case of an anomalous observation, i.e. the situation that E contradicts B, the task is to revise B to B^* such that B^* entails E. This is called *anomaly guided abduction*. There are at least two different types of revision in this case. One is similar to the so-called belief revision program: first contract B to some weaker B', then expand B' with some H such that B'&H&E are consistent, neither B' nor H entails E, and B'&H (=B^*) entails E. The other type of revision, presumably dominating in scientific practice, consists of 'concretization' of B to B^*, where B can be seen as an 'idealization' of B^*. This second type of revision falls as yet outside the reach of (most of) the algorithms considered. Hence, we focus on the first, contraction-followed-by-expansion, type of revision.

To call not only the first but also the second, anomaly guided, abduction task 'standard', is only defensible as far as approaches are concerned that connect abduction with belief revision or with applications in philosophy of science. Logical AI approaches and abductive logic programming, for example, only deal with the first, novelty guided, abduction task.

The main point of this paper is to generalize the standard abduction tasks to theory revision aiming at empirical progress (Task III) and truth approximation (Task IV). Here, 'theory revision' may or may not be conceived by definition as 'theory revision within a certain research program'. Below we will further explicate these tasks in such a way that the following claims can be argued for. The first claim is that Task I and II are special cases of Task III. The second claim is that Task IV amounts to Task III and some (additional) evaluation tasks. Hence, Task III or, more precisely, the subtask called *instrumentalist abduction* can be viewed as the main abduction task. The paper concludes with a survey of possible points of departure, in AI and logic, for computational treatment of the instrumentalist task guided by the 'comparative evaluation matrix'.

2. ANALYSIS OF THE EMPIRICAL PROGRESS TASK

Task III, that is, theory revision aiming at empirical progress, requires first of all an explication of the idea that a theory Y is an empirical improvement over theory X. Our point of departure for this explication is the *evaluation report* of a theory in terms of its successes, failures and lacunae, according to the data at a certain time. More specifically, *successes* of a theory are conceived as explanatory successes, that is, observations that can be explained by the theory, which include observations that have been successfully predicted. Here, 'explanation' is used in the liberal sense of deductive or some other type of entailment of the relevant statement by the theory, without assuming that the theory is true, or at least not known to be false. On the contrary, from the present point of view, the truth or falsity of a theory is not so important. What matters is the success domination (see below) when compared with other theories. This is directly related to the truth approximation point of view: a false theory may well be very close to the truth (in the sense to be specified), and if it is closer to the truth than another one, it will be at least as successful. The *failures* of a theory amount to so-called anomalous observations or counterexamples, the descriptions of which contradict the theory. Finally, *lacunae* pertain to surprising observations, that is, observations that cannot be explained by the theory, let alone predicted, but are at the same time not in conflict with the theory.

Below we will define the idea that theory Y is more successful than X at t in terms of their respective evaluation reports. Presupposing such a definition, Task III can be divided into two subtasks:

Task III search for a revision Y of X such that

1 Y is more successful than X, relative to the available data, and

2 Y remains more successful than X, relative to all future data

Whereas Task III.1 is really a subtask of Task III, Task III.2 essentially is a strengthening of the former.

Task III.1 may be called *the instrumentalist task* since ‘more successful’ is, of course, defined in observation terms (relative to the background knowledge) and Y may well be ‘born refuted’, that is, Y may have failures. If so, they should be (at least as serious) failures of X as well.

Clause 2 amounts to a further condition on Y relative to X. Suppose we have constructed a theory Y that is, relative to the available data, more successful than X. Then the question remains whether it will remain at least as successful as X relative to all future data. This amounts to what might be called the empirical progress hypothesis or *comparative success hypothesis* (CSH). It is important to realize that CSH is a decent empirical, that is, falsifiable, hypothesis, which is serious even if Y is known to be false. CSH can be tested by suitable experiments and will be accepted, for the time being, when ‘sufficiently many and varied experiments’ have failed to falsify it. In that case, empirical progress has been made, or at least it seems so.

3. ANALYSIS OF THE TRUTH APPROXIMATION TASK

Recall that Task IV dealt with theory revision aiming at truth approximation:

Task IV search for a revision Y of X such that Y is closer to the truth than X

The condition will be called the *truth approximation hypothesis* (TAH). Note again that TAH is serious even if Y is (known to be) false. To be sure, the condition presupposes a definition of 'closer to the truth' and hence, first of all, a definition of 'the truth'. The latter is defined relative to the (union of) vocabularies of X and Y and the relevant domain as the strongest, hence most informative, true hypothesis that can be formulated within the vocabulary about the domain. Hence, 'the truth' is conceived in a constructive, non-essentialistic way. In the natural sciences the domain pertains to all physical possibilities of a certain kind, in which case the truth will not be complete (as long as we do not include modal operators in the vocabulary). Assuming this modal type of 'the truth', it is possible to define 'closer to the truth' (or 'more truthlikeness') in a naive and refined way such that everybody will at least agree that the defining condition is a sufficient one. Unfortunately, it would lead too far afield to elaborate these definitions here in a formal sense, not least because one can give three versions of increasing strength. For this purpose, the reader is referred to (Kuipers, 1997, 1999) for details. Here it will be sufficient to give verbal formulations of the three (naive) versions, in the context of a further analysis of Task IV.

Task IV search for a revision Y of X such that Y is

1 closer to the observational truth than X and

2 closer to the referential truth than X and

3 closer to the theoretical truth than X

In Task IV.1 the truth is restricted to the observational vocabulary. In the next section we will argue in some detail that the first clause amounts to the conjunction of clauses 1 and 2 of Task III. For this reason it will be called *the (constructive) empiricist task*, where we freely, but plausibly, presuppose a modal version of (constructive) empiricism.

Clause 2 of Task IV adds the requirement that Y should be referentially better than X in the sense that the referential claims of Y regarding the joint theoretical vocabulary are at least as adequate as those of X and superior in at least one respect. Here ‘referential claims’ are interpreted as ‘contributing to the restriction of models of the theory or not’, and they are true when this is so according to ‘the referential truth’, that is, the truth regarding the reference of the theoretical terms, and false otherwise.

Finally, clause 3, to be called *the constructive realist task*, adds that Y should be closer to the theoretical truth. Of course, the theoretical truth is defined in such a way that it implies the observational as well as the referential truth in the sense that it reduces to these truths when we restrict the attention to the observational vocabulary and to referential claims of the theoretical vocabulary, respectively.

The general idea of the (naive, qualitative) truth approximation theory is that Y is at least as close to the truth as X when Y shares all correct models of X and X shares all incorrect models of Y, where a (in)correct model is of course a model that does (not) belong to the (models of the) truth, that is, the strongest true theory. This idea leaves room for theoretical and observational specifications in the case of a stratified vocabulary.

4. RELATIONS BETWEEN EMPIRICAL PROGRESS AND (KINDS OF) TRUTH APPROXIMATION

In (Kuipers, 1999) we have extensively argued that there are strong relations between the various kinds of truth and successfulness and hence between the corresponding abduction tasks. To begin with, Task IV.3 is likely to be fulfilled when Task IV.2 is, but it is not guaranteed. Moreover, Task IV.2 may well be fulfilled when Task IV.1 is, but it is far from likely. At most one might say that it is more likely than not. Hence, when Task IV.1 is fulfilled, additional testing of the referential truth approximation hypothesis is required.

When and only when evidence from different directions converges (a happy situation called triangulation of evidence) do scientists tend to accept referential claims.

We noted already that Task IV.1 essentially amounts to Task III.2. In one direction it is almost obvious in the following sense: 'closer to the truth', when construed in a safe way, not only implies 'at least as successful' but even the possibility of some extra success. But it is also plausible in the other direction. For 'remaining more successful' could be violated when Y is 'not closer to the observational truth' than X, since the latter would leave room for an extra success of X relative to Y.

So let us turn to Task III.2, assuming that Task III.1 has delivered a suitable Y. The second clause requires further comparative testing of Y against X. This asks for the generation and testing of distinctive test implications. Relative to Task III.1, Task III.2 amounts to the evaluation phase of a hypothesis generated by the former. Hence, there remains as the abduction task in the strict sense, Task III.1, the instrumentalist task.

In sum, abduction aiming at empirical progress amounts to the generation Task III.1 and the further evaluation task required by Task III.2. Moreover, abduction aiming at truth approximation requires in *addition* only some further evaluation tasks, in particular regarding reference claims. To be sure, selection is not yet included in this picture, but it is fairly obvious how that could be done.

Suppose that the generation task has produced two alternative revisions of X, viz. Y and Z, and suppose that both survive the additional evaluation tasks. Then the selection question is of course whether one is a suitable revision of the other. To begin with, e.g. is Z more successful than Y, relative to the available data? If so, the other evaluation questions have to be raised. If not, and if Y also fails to be more successful than Z, that is, if Y and Z have 'divided success', there arises a new generation task, viz. to construct a hypothesis that is more successful than Y and Z, and hence, than X. Of course, this selection story can be generalized to finitely many theories generated as improvements of X. The conclusion remains that the primary abduction task aiming at empirical progress or even at truth approximation is the instrumentalist task III.1. The rest is evaluation and selection, and possibly new generation tasks for further improvement.

5. THE REMAINING INSTRUMENTALIST TASK

Recall that Task III.1 amounts to:

> search for a revision Y of X (possibly within one research program) such that Y is more successful than X at time t

For the crucial definition of 'more successful' there are two related alternatives, an asymmetric and a symmetric one. Both definitions may be called 'dominance' definitions, providing at least sufficient conditions, and hence leaving room for more liberal definitions. Moreover, in both cases there are naive and refined versions, but we restrict the attention to the naive versions. The asymmetric one reads:

Y is (at t) *more successful than* X iff (at t)

- X shares Y's individual counterexamples
- Y shares X's general explanatory successes
- at least once, not vice versa

This definition seems in accordance with scientific practice and is in close harmony with a similar, intuitively appealing, asymmetric definition of 'closer to the truth'. However, there are also two symmetric definitions, an individual and a general one, that are at least as much in accordance with scientific practice, but less straightforwardly related to the intuitively appealing definition of 'closer to the truth'. The symmetric definitions are in terms of individual or general successes, lacunae and failures and are suggested by the following comparative evaluation matrix for individual or general data:

		X		
		Failures	Lacunae	Successes
	Failures	0	–	–
Y	Lacunae	+	0	–
	Successes	+	+	0

The (comparative) evaluation matrix

Besides neutral results (0), there are three types of results which are favorable (+) for Y relative to X, and three types of results which are unfavorable (−) for Y relative to X. Now Y is more successful than X when there are, besides neutral results, some favorable results for Y and no unfavorable results for Y.

It is not difficult to check that both symmetric definitions and the asymmetric one are essentially equivalent, that is, equivalent assuming some boundary conditions. Moreover, in particular in the symmetric set-up, it is clear that the different types of favorable results could be given different weights, which may be useful for non-dominance definitions. Another point of refinement is that failures, lacunae and successes may be given different weights even if they belong to the same comparative category. E.g. as Hanne Andersen (to appear) notes, anomalies may be graded as 'mere' anomalies, important anomalies or even severe ones.

Finally, it is clear from the symmetric definition that Task I and II are special cases of Task III.1. In the case of Task I, dealing with a surprising event, we have an individual or general fact which is a lacuna for X, but a(n) (explanatory) success of Y. In the case of Task II, dealing with an anomalous event, we have an individual or general fact which is a failure for X, and a success of Y. Hence, the two standard abduction tasks are included in the instrumentalist abduction task. In other words, the latter is a generalization of the standard tasks. However, in the two standard cases the revised theory has to be compatible with all the available data, whereas this need no longer be the case in the non-standard cases. This is one of the fundamental differences, if not the most fundamental one, between standard and non-standard cases.

The resulting invitation for abduction aiming at empirical progress and even truth approximation reads: design instrumentalist abduction along symmetric or asymmetric lines. At first sight, it may seem to be a generalization of the standard tasks with fewer constraints, since no consistency between the new theory and the data is required; hence the threat of becoming intractable seems real. However, since the required piecemeal improvements need to avoid drawbacks, every success and every lacuna (not being the target) provide an extra constraint.

6. AI-PROSPECTS FOR COMPUTATIONAL APPROACHES TO INSTRUMENTALIST ABDUCTION

There are several computational ideas and computer programs dealing with (standard) abduction and theory revision. Here we will give an (incomplete) survey of them. It is only meant as a starter for a research project that first tries to complete the present survey as much as possible and next selects one or two possibilities for computational elaboration in the direction of the above described instrumentalist abduction task. At first sight, all indicated programs seem worthy of further investigation for this purpose, in particular when guided by the (symmetric) comparative evaluation matrix, for in most programs the data to be explained and the anomalous data are of the same, individual or general, nature.

We start by reviewing some of the leading AI-ideas/programs for abduction and theory revision, viz. Thagard's program PI and those presented in (Shrager and Langley, 1990). In Section 7 we will deal with some logic oriented approaches. In the final section we will say a few words about evaluation and selection.

Paul Thagard's program PI (Processes of Induction) (Thagard, 1988) is based on a distinction between six forms of induction, four of which are called abduction: individual, rule, existential, and analogical abduction. All of them are basically cases of standard novelty guided abduction. PI is grafted upon models of heuristic search developed in cognitive psychology. Hence, abductive explanatory problems are approached by PI, in a quasi-connectionist way, by activation of the concepts occurring in the problem, followed by the activation of concepts coupled to the first ones by rules ('rule firing'). PI does not deal with anomaly guided abduction, and, more generally, in its present form PI has serious limitations in the sphere of evaluation and selection, since it is restricted to still unfalsified theories (see below). However, it seems certainly worthwhile to try to revise PI in order to overcome these limitations (for some suggestions, see Kuipers, 1993).

Let us now turn to AI-programs for abduction and theory revision presented in (Shrager and Langley, 1990). The 'Abduction Engine' (AbE) of O'Rorke, Morris and Schulenburg pretends to make revolutionary revision of a theory possible, starting from an observational anomaly between the theory, the knowledge base, and new data. AbE uses Forbus's Qualitative Process language (Forbus, 1984) for describing qualitative changes due to processes acting on quantities. After the recognition of an anomaly, the theory, conceived as a hierarchically ordered set of principles, is weakened (contraction) to a basic theory that is compatible with the new data. Next, this basic theory is strengthened (expansion), roughly along similar lines as in Thagard's PI, to a theory that is able to explain the anomalous data. The program is developed on the basis of the transition of the phlogiston theory to the oxygen theory of combustion. Hence, it is no surprise that the program is able to reproduce that episode. It is easy to recognize in this step-wise procedure an overall case of anomaly guided abduction, and a subcase of novelty guided abduction. Hence, it deals with standard cases of instrumentalist abduction.

In the program COAST of Rajamoney, also using Forbus's Qualitative Process language, theory revision is approached step-wise: 1) registration of an anomaly between new data and a theory in the knowledge base, 2) design of revision proposals that transform the anomalous data into explanatory successes, where revision operators act on components, domains and effects, 3) suggestion of (new) experiments, 4) first selection on the basis of the anomaly to be explained, the outcome of the experiments, and the degree in which previous successes are retained, 5) further selection on the basis of simplicity and predictive power (strength). It is easy to recognize in these steps standard abductive steps and consecutive evaluation and selection phases. COAST, and the following two programs, were guided in their development by examples from biochemistry.

The program KEKADA of Kulkarni and Simon is globally similar to COAST, apart from differences indicated below. It registers surprising phenomena by comparing new experimental results, obtained for 'normal' purposes, with expectations on the basis of the knowledge base, and proceeds by dealing with such phenomena. That is, it tries to revise the relevant theory and proposes experiments to test such revisions. For this purpose, five strategies are built into the program, each of which consists of an hypothesis generator, an experiment proposer and an evaluator. For example, the core idea of the first strategy is to try to strengthen the surprising phenomenon by independently manipulating the variables of the system.

Karp's program HYPGENE is conceptually rather different from the previous ones. It considers hypothesis formation and revision explicitly as a matter of design, with constraints (a profile of desired properties) and operators that revise the provisional prototype (the profile of factual properties) in order to let it satisfy the constraints better and better. Here for the first time, non-standard cases are naturally included for the revised prototype may well have failures. As was pointed out in (Kuipers, Vos and Sie, 1992), the symmetric definitions of 'more successful' (and of 'closer to the truth') are completely analogous to the claim that one prototype is an improvement of another, relative to a fixed intended profile. In other words, the (ultimately) intended profile, in case of instrumentalist abduction, is a theory for which all data are successes. Van den Bosch (1999) shows in this volume that a similar strategy can be based on Anderson's ACT-R.

Finally, Darden prepares in the volume of Shrager and Langley the conceptual ground for a program still to be written by considering the solution of an anomaly of a theory as a task for diagnostic reasoning, which has been developed for expert systems, guided by the tracing of a defect in a technical system. She shows with Mendelian examples (see also Darden, 1991) that the hierarchical decomposition of all presuppositions of a theory may provide the points of departure for solving the anomaly, and that its solution, as in the famous analysis of some mathematical examples by Lakatos (1976), may or may not lead to fundamental theory revision, that is, a revision staying within a research program or breaking through the barriers of the relevant program.

7. LOGICAL PROSPECTS FOR COMPUTATIONAL APPROACHES TO INSTRUMENTALIST ABDUCTION

One challenge for logic is the question to what extent the AI-approaches sketched in the previous section can be reproduced in a logically decent way. However, we will not pursue this further here, and will concentrate on already existing purely or primarily logical approaches. As far as we know, there are only a few logical approaches to standard abduction and theory revision, viz. so-called belief revision developed by Peter Gärdenfors and others (Gärdenfors, 1988), non-monotonic logic approaches by Konolige (1990, 1996) and the semantic tableau approach introduced by Mayer and Pirri (1993) (see also Pirri, 1995), which was recently redirected in a more efficient way by Atocha Aliseda (1997). The precise relation between belief revision and non-monotonic logic approaches on the one hand and instrumentalist abduction aiming at empirical progress and truth approximation on the other, has still to be investigated. For example, *prima facie*, there are some important limitations of belief revision, as seen from the instrumentalist abduction task. To begin with, belief revision is rather 'actual world' oriented, instead of oriented toward 'physically possible worlds' or, simply, 'real world' oriented. Moreover, it aims at a consistent product of the available data and the resulting belief set. The latter feature it shares, however, with almost all approaches, including the one by Aliseda.

Let us turn to a brief indication of Aliseda's semantic tableau method of novelty and anomaly guided abduction. As remarked already earlier, her conceptual approach guided our presentation in several respects. Here we only want to indicate briefly the basic idea behind her algorithms, which are as yet restricted to propositional languages. Recall that a semantic tableau is a way to find out whether a set of premises logically entails a conclusion by systematically trying to construct a counterexample, that is, a propositional model or structure in which the premises are true and the conclusion is false. If all branches generated by the (formula decomposing) tableau rules from the set of premises and the negated conclusion become 'closed', that is, contain a formula and its negation, the argument is valid. Open branches describe counterexamples to the argument. In the case of novelty guided abduction, there are of course open branches in the tableau starting from the theory and the negation of the phenomenon description. Systematically closing these branches produces equally many consistent proposals for abductive explanation of the novel phenomenon. In the case of anomaly guided abduction, the theory apparently entails the negation of the anomaly, that is, the corresponding tableau starting from the theory and the phenomenon description is closed. Now the task is to systematically open the branches of this tableau by deleting parts of the theory (contraction), followed by adding new formulas (expansion) that lead to the closing of all branches when combined with the negation of the phenomenon description.

The above description neglects all nuances and versions of Aliseda's approach, but it may nevertheless already be plausible to conjecture that the approach can be generalized to instrumentalist abduction purposes. The main thing is to operationalize in tableau terms the possibility that one theory is 'more incompatible' with the data than another. Then the task is to generate theories that are not more incompatible with the available data than the original one, explain at least as much as the latter, and improve the latter in at least one of these respects. Some other adaptations will be required to make it suitable for scientific purposes. First, there should be left room for general facts as data and, second, the original and resulting theories should be of a general nature. When restricted to propositional languages, simple implicative formulas will do the job in both cases, provided there is left room for three kinds of truth: actually true, 'physically true' and logically true. One plausible way to do this is using a modal propositional language (Zwart, 1988), but there may well be other ways.

Let me conclude this section with some specific remarks. The first deals with theory structure. In several approaches there is a natural tendency to conceive theories in a hierarchical way (notably O'Rorke c.c. and Darden). Starting from intuitive hierarchies as used by scientists, this is easy to reproduce in syntactic as well as semantic or structuralist axiomatizations of theories. However, when such intuitive hierarchies are not available, the method of 'natural axiomatization' developed by Gemes (1994) may well be helpful in producing hierarchies.

The second remark pertains to the restriction to the 'naive' instrumentalist abduction task, viz. all examples of theory revision presented so far deal with the three possible relations between a theory and one individual or general fact: entailment, incompatibility, and no logical relation. The 'refined' instrumentalist abduction task asks for refinement of the second and third alternative: one theory may approach or explain a (possibly incompatible) fact better than another. It is likely that programs aiming at quantitative theory generation and revision are more appropriate for these purposes. As a matter of fact, it is plausible to conjecture that the BACON-programs, developed by Simon and others, can be reconstructed as refined revision programs of quantitative (quasi-true, that is, at most approximately true) laws. Although it will be worthwhile to pursue this line further, and also to develop something like refined theory revision in the sense of 'concretization' of quantitative theories, the naive instrumentalist abduction task is at least as important since it deals with qualitative theories. Before refining propositional instrumentalist abduction, it will be necessary to fully develop a naive version. Moreover, the generalization of the suggested naive propositional instrumentalist abduction in tableau terms, to first order logic may be considered as at least as important as the indicated refinement.

8. CONCLUDING REMARKS

Sections 5–7 may be seen as a description of a research project aiming at the computational realization of instrumentalist abduction. We close by indicating some general features of such programs. First, although the terminology of theory revision suggests that there should be a contingent initial theory, this is not so. It is possible to start from the theory representing *ignorance*, that is, a tautology, in which case the task is to improve upon ignorance. Second, one may or may not aim at *incremental* abductive programs. A program is incremental, starting from a certain theory and a certain set of data, when the revised version dealing with these data, is the point of departure when the set of data further increases. The non-incremental alternative is that the second revision process starts again from scratch, that is, from the original theory. Third, when the generation process generates alternative theories, of which one is more successful than all the other ones, the selection is clear. Depending on the further aims of the program, the selection may then be followed by further evaluation subprograms dealing with empirical progress or truth approximation. When the generated theories cannot be ordered in the suggested way due to 'divided success' between any pair of theories, the generation task has not yet been completed. Finally, when the result contains a couple of theories that are at least as successful as all other theories, one may or may not want to have further selection. However, as long as the data remain the same, one will have to use non-empirical features, such as *simplicity* and *strength*. The strength criterion is straightforwardly functional for empirical progress and hence for truth approximation, in the sense that new data may well lead to the conclusion that a stronger theory is more successful than a weaker one. However, for the simplicity criterion such straightforward functionality does not seem to exist as long as it does not imply the strength criterion. On the contrary, as McAllister (1996) argues and illustrates, simplicity or, more generally, beauty considerations may well retard empirical progress.

ACKNOWLEDGEMENTS

I would like to thank Atocha Aliseda, Johan van Benthem, Alexander van den Bosch, Hidde de Jong and the editors for their helpful and interesting comments.

REFERENCES

A. Aliseda: 1997, *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence*. Dissertation Stanford, ILLC Dissertations Series (1997-04), Amsterdam.

H. Andersen: to appear, Unexpected Discoveries, Graded Structures, and the Difference between Acceptance and Neglection. In J. Meheus and T. Nickles (eds.), *Scientific Discovery and Creativity*. Kluwer, Dordrecht.

A. van den Bosch: 1999, Qualitative Drug Lead Discovery. *Foundations of Science* 4(4) (to appear).

L. Darden: 1991, *Theory Change in Science: Strategies from Mendelian Genetics*. Oxford UP, Oxford.

K. Forbus: 1984, Qualitative Process Theory. *Artificial Intelligence* 24: 85–186.

P. Gärdenfors: 1988, *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA.

K. Gemes: 1994, A New Theory of Content. *Journal of Philosophical Logic* 23: 596–620.

K. Konolige: 1990, A General Theory of Abduction. In *Automated Abduction*, Working Notes. Spring Symposium Series of the AAA. Stanford University, Stanford, pp. 62–66.

K. Konolige: 1996, Abductive Theories in Artificial Intelligence. In G. Brewka (ed.), *Principles of Knowledge Representations*. CSLI Publications. Stanford University, Stanford.

T. Kuipers: 1993, On the Architecture of Computational Theory Selection. In R. Casati and G. White (eds.), *Philosophy and the Cognitive Sciences*. Austrian Ludwig Wittgenstein Society, Kirchberg, pp. 271–278.

T. Kuipers: 1997, The Dual Foundation of Qualitative Truth Approximation. *Erkenntnis* 47.2: 145–179.

T. Kuipers: 1999, *From Instrumentalism to Constructive Realism. On some Relations between Confirmation, Empirical Progress and Truth Approximation*. Manuscript, to appear in 2000 in the Synthese Library, Kluwer, Dordrecht.

T. Kuipers, R. Vos and H. Sie: 1992, Design Research Programs and the Logic of their Development. *Erkenntnis* 37.1: 37–63.

I. Lakatos: 1976, *Proofs and Refutations*. Cambridge University Press, Cambridge.

M. Mayer and F. Pirri: 1993, First Order Abduction via Tableau and Sequent Calculi. *Bulletin of the IGPL* 1: 99–117.

J. McAllister: 1996, *Beauty and Revolution in Science*. Cornell UP, Ithaca.

F. Pirri: 1995, *Abduction*. PhD Dissertation, Université de Paris VI, Pierre et Marie Curie, Paris.

J. Shrager and P. Langley (eds.): 1990, *Computational Models of Scientific Discovery and Theory Formation*. Kaufmann, San Mateo.

P. Thagard: 1988, *Computational Philosophy of Science*. MIT Press, Cambridge.

S. Zwart: 1988, *Approach to the Truth. Verisimilitude and Truthlikeness*. Dissertation Groningen, ILLC Dissertations Series (1998-02), Amsterdam. A revised version will appear in the Synthese Library of Kluwer AP, Dordrecht.

Faculty of Philosophy
University of Groningen
T.A.F.Kuipers@philos.rug.nl